

Refining disparity maps using deep learning and edge-aware smoothing filter

Shamsul Fakhar Abd Gani^{1,2}, Muhammad Fahmi Miskon¹, Rostam Affendi Hamzah², Mohd Saad Hamid², Ahmad Fauzan Kadmin², Adi Irwan Herman³

¹Fakulti Kejuruteraan Elektrik, Universiti Teknikal Malaysia Melaka, Malaysia

²Fakulti Teknologi Kejuruteraan Elektrik dan Elektronik, Universiti Teknikal Malaysia Melaka, Malaysia

³Product and Test Engineering, Texas Instruments (Malaysia), Melaka, Malaysia

Article Info

Article history:

Received Apr 14, 2023

Revised Sep 29, 2023

Accepted Dec 21, 2023

Keywords:

Computer vision

Deep learning

Disparity map

Edge-aware filter

Stereo matching

ABSTRACT

Stereo matching algorithm is crucial for applications that rely on three-dimensional (3D) surface reconstruction, producing a disparity map that contains depth information by computing the disparity values between corresponding points from a stereo image pair. In order to yield desirable results, the proposed stereo matching algorithm must possess a high degree of resilience against radiometric variation and edge inconsistencies. In this article convolutional neural network (CNN) is employed in the first stage to generate the raw matching cost, which is subsequently filtered with a bilateral filter (BF) and applied with cross-based cost aggregation (CBCA) during the cost aggregation stage to enhance precision. Winner-take-all (WTA) strategy is implemented to normalise the disparity map values. Finally, the resulting output is subjected to an edge-aware smoothing filter (EASF) to reduce the noise. Due to its resistance to high contrast and brightness, the filter is found to be effective in refining and eliminating noise from the output image. Despite discontinuities like adiron's lost cup handle or artl's shattered rods, this approach, based on experimental research utilizing a Middlebury standard validation benchmark, yields a high level of accuracy, with an average non-occluded error of 6.79%, comparable to other published methods.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Shamsul Fakhar Abd Gani

Fakulti Teknologi Kejuruteraan Elektrik dan Elektronik

Universiti Teknikal Malaysia Melaka

Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

Email: shamsulfakhar@utem.edu.my

1. INTRODUCTION

Light detection and ranging (LIDAR) is a remote sensing approach that employs a pulsed laser to detect distances and build three-dimensional (3D) discrete surfaces of the real world. Stereo vision addresses LIDAR's hardware, interface, and power issues using left and right cameras to take images of the same target from various angles and match pixels to calculate depth. Stereo vision estimates depth via triangulation. By knowing the distance between the cameras and the angle between their optical axes, the depth of any point in the scene that matches in both images can be found by using trigonometry as shown in (1):

$$z = \frac{f \cdot b}{d} \quad (1)$$

where z is the depth, f is the camera focal length, b is the baseline space in the middle of cameras' optical centre, with d as disparity. Stereo vision has many applications in fields such as robotics, 3D object/face recognition [1]-[7], and virtual reality. Stereo vision struggles with occlusion, illumination, noise, calibration errors, and computing complexity. Many methods and approaches have been developed to increase stereo vision system accuracy and efficiency. Scharstein *et al.* [8] suggested a common outline which was executed by utilizing a sequenced multi-stage system shown in Figure 1. Rectified image pairs are fed to the framework. Then, calculate the cost function to quantify patch similarity in the input image pair. The second step aggregates costs and filters noise [9]-[11]. Then, winner-take-all (WTA) strategy is employed to select the disparity with the lowest cost while discarding the others. Finally, disparity refinement is achieved by applying optical low-pass filter among other refinement methods.

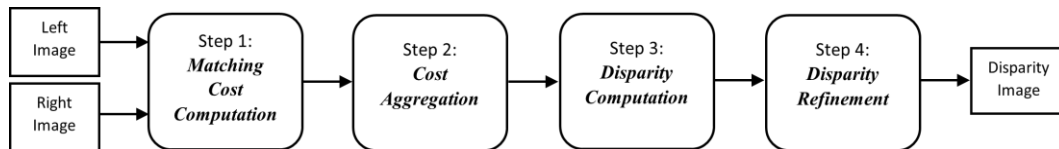


Figure 1. The stages of the stereo vision framework [8]

Local approaches utilise disparity based on the relationship between pixel intensities (grayscale, RGB colours, texture patterns) inside a certain local support window based on the connections between pixels in close proximity to one another in the corresponding image [12]. These methods are sometimes referred to as window-based or region-based approaches. Locally established techniques include sum of absolute differences (SAD) [13], sum of squared differences (SSD) [14], and normalised cross-correlation (NCC) [15]. Local stereo matching techniques estimate the disparity by comparing the surroundings of a pixel p in the left image to the surroundings of a pixel q in the right image, where q has been translated across a potential disparity p . Local method evaluation leads in a quick runtime, but unfortunately, the quality is compromised, particularly in the region of depth discontinuities.

Global optimization methods on the other hand address the disparity problem as a reduction of a predetermined global energy function [16]. It typically requires additional resources and is less susceptible to local variance. The evaluation is computed using global data and a smoothing threshold for neighbouring pixels. Markov random field (MRF) has generated a variety of answers to the problem of global energy reductions. These approaches can either be classified as graph cut (GC) or belief propagation (BP) methods. GC generates the lowest energy solution by integrating the smallest cutoff and maximum flow of the retrieved MRF graph. In contrast, the BP technique reduces the energy function by continually transmitting signals from the present node to neighbouring nodes within the MRF network [17]-[20].

Since [21] introduction of convolutional neural network (CNN) trained on pairs of tiny image patches with known real disparity, interest in a deep learning-based stereo vision system has increased substantially. Some post-processing tweaks are still required for these approaches. Several deep-learning based stereo matching algorithms, like DispNets [22], GCNet [23], PSMNet [24], Gwc [25], EdgeStereo [26], LEAStereo [27], HITNet [28], and EAI-Stereo [29] minimise post-processing stages and improves stereo matching performance.

2. METHOD

Figure 2 is a block diagram of the proposed method for the experiment, with the 4 boxes representing the adapted 4 stages from Figure 1. Edge-aware smoothing filter (EASF) [30] is performed to remove noise created from filling in process.

- Matching cost computation: the stereo matching method starts by extracting the height, width, and number of disparities from the calibration file. The left and right images are then loaded using the OpenCV library, and pre-processing techniques like scaling, grayscale conversion, and normalisation are applied. The image is then processed by a trained CNN model implemented with tensorflow to obtain the cost volume, a 3D matrix containing the matching cost for each pixel and disparity level.

As shown in Figure 3, a series of square-shaped kernels make up the convolution layer, and despite their small size, these filters can fit the whole depth of the volume. Each kernel reads the input region, adds the dot product, and stores the result in the activation layer. The input volume for the succeeding network layer is then constructed by integrating the feature map layer. The computing cost of adding more rectified

linear unit (ReLU) layers to a CNN grows linearly [31], [32]. Pooling layer gradually reduces the input dimension, reducing system parameters and processing, and aiding in the prevention of fitting problems.

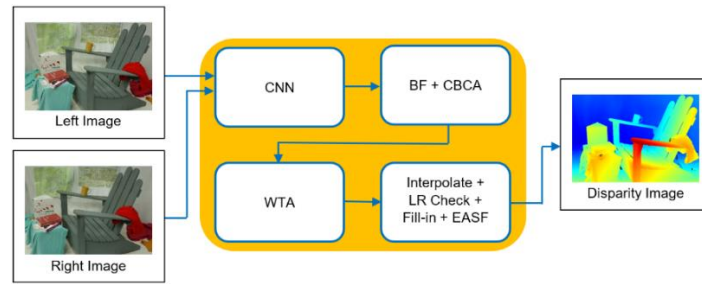


Figure 2. The proposed method

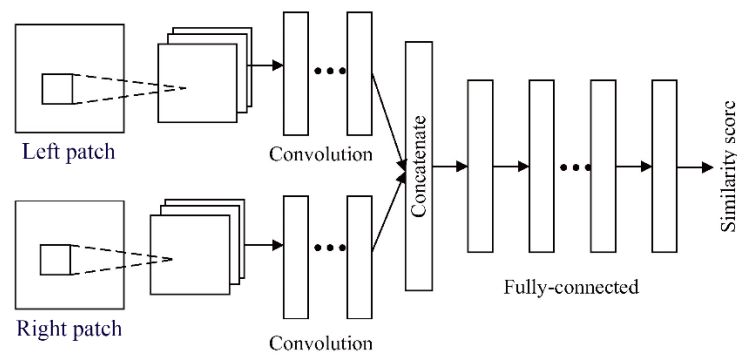


Figure 3. The structure of MC-CNN [13]

The cost of matching may therefore be simply computed based on the CNN output. Based on [25], C_{CNN} reflects the cost value as (2):

$$C_{CNN}(p, d) = -s(< P^L(p), P^R(p - d) >) \quad (2)$$

where P is the patch for L (left) and R (right), p is the (x, y) position, and d is the disparity. The output of layers $L1$, $L2$, and $L3$ need to be computed only once per location p and need not be recomputed for every disparity d . Using a patch-based technique, features for both the left and right images are computed, utilising the left and right images as inputs along with the patch height and patch width. From there, the cost volume is computed by taking the left and right feature matrices along with the expected maximum disparity ($ndisp$) and calculating the matching cost for each potential disparity level by comparing the left and right features and iterating over all possible disparities, which range from 0 to $ndisp-1$. For each disparity, the function calculates the dot product of the features at each pixel in the left image and the corresponding pixel in the right image shifted by the disparity.

- Cost aggregation: in cost aggregation stage, cost filtering is first done using BF which is useful for removing noise while preserving edges [33]. The BF takes in three arguments: the input image, the size of the filter kernel (diameter of each pixel neighborhood), and two parameters controlling the filter's range of influence: sigma color and sigma space. Sigma color controls how much the pixel values can differ while still being considered neighbors, and sigma space controls how far away the pixels can be while still being considered neighbors. The function loops over each disparity level d and applies the bilateral filter to the 2D slice of the cost volume at that disparity level. The filtered slice is then stored in the corresponding slice of a new 3D array with the same dimensions as the raw cost volume.

Cross-based cost aggregation (CBCA) shown in Figure 4 [34] is applied to the filtered cost volume to further refine the disparity map by computing the cross region of the image. The cross region is constructed into vertical and horizontal regions for efficiency, and the function returns a numpy array representing the cross union region and a numpy array representing the number of elements in each position of the cross union region. CBCA aggregates the matching costs of neighboring pixels and disparities to reduce errors caused by occlusions and noise. First the disparity map is initialized with the minimum cost disparity for each pixel. Then, for each pixel, the algorithm selects the disparity with the minimum cost and

applies a weighted median filter to the neighboring pixels and disparities that have a similar cost. This process is repeated for several iterations to refine the disparity map further.

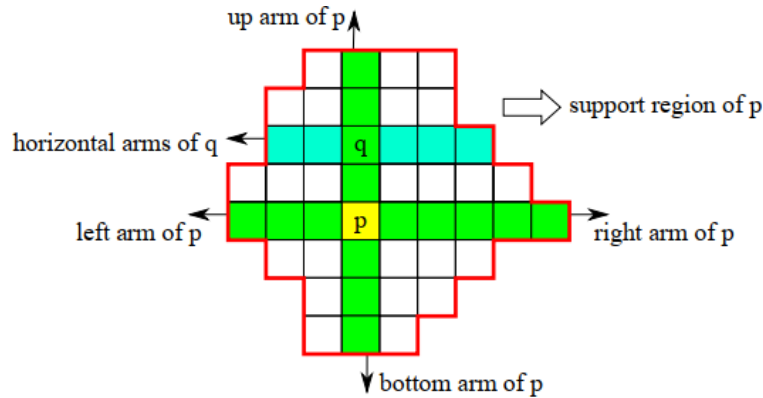


Figure 4. Cross-based cost aggregation

In (3) shows the BF function used in this paper.

$$BF[I]_p = \frac{1}{w_p} \sum_{q \in s} G_{\sigma s}(|p - q|) G_{\sigma r}(CNN_p - CNN_q) I_q \quad (3)$$

where $G_{\sigma s}$ is ‘space’ parameter which determines the positive effect of faint pixels, $G_{\sigma r}$ is ‘range’ parameter which determines the impact of pixel q having a concentration amount varying from I_p .

- Disparity computation: disparity computation determines the specific combination of disparities for which normalisation of disparity values occurs. In this stage, the most common local technique, WTA optimization, is implemented which is shown in (4):

$$d_p = \arg \min_{d \in D} C(BF[I]_p, d) \quad (4)$$

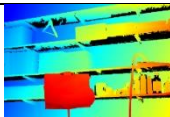
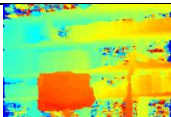
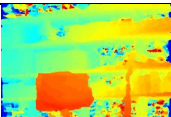

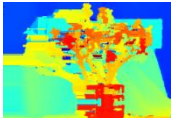
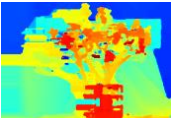
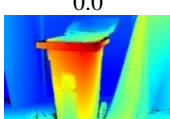
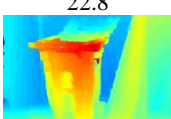
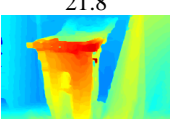
where d_p is the disparity with the least expensive quantity selected, $C(BF[I]_p, d)$ is the cumulative cost applied with CBCA and BF in the second stage, D is the set of all acceptable individual disparity. The index of the minimum matching cost is obtained using the `np.argmin()` function and assigned as the label for the pixel in the respective label map. WTA labels each pixel with the disparity with the lowest matching cost without considering neighbouring pixels, which can produce noisy results in textureless or occluded regions.

- Disparity refinement: in the last stage of the framework, standard approaches for post-processing and disparity refinement were implemented. Peng *et al.* [35] integrated the mean shift and super-pixel with segmentation (SEG) approach which clusters the disparity map according to colour.
- Weighted median filter (WMF) is implemented where it essentially combines box aggregation with a weighted median, effectively eliminates noise from outliers while conserving the edges [36]. Median filter (MF) was utilised in [35]. MF is accurate at the edges but inaccurate in low-textured regions. BF was used to enhance the edge quality, despite the fact that its computation took longer than that of other filters [37]. Disparity maps may have missing pixels due to occlusions, depth discontinuities, or noise. Interpolation fills these gaps using left-right consistency check and median filtering. Mismatches are interpolated with median value of nearest matching neighbors, while occlusions use disparity value of nearest matching neighbor on right (or left) or raw disparity value if no match found. The quality of the disparity map is enhanced further through the application of EASF which is a modification of the traditional edge-preserving filter [38], [39] that iteratively applies the filter, smoothing the image while preserving its edges [40]. The idea is that the amount of smoothing between two pixels should depend on how far apart they are. The further apart they are, the less smoothing should occur between them. This can be expressed mathematically using a distance metric in a 5D space. If a transformation is applied that preserves these distances, then the edge-preserving property of the filter will also be maintained even if the image is represented in a lower-dimensional space. The recursive nature of this filter allows it to better preserve the edges in an image while still removing noise. This is because the filter is able to refine its output over multiple iterations, gradually removing noise while preserving the important features of the image.

3. RESULTS AND DISCUSSION

Experiment is run on Windows 10 computer with Intel Xeon 2.80 GHz CPU, Nvidia Quadro P1000, and 8 GB DDR4 RAM. Project is written in python, utilizing libraries such as tensorflow, pytorch, and keras, making it easy to construct and train deep learning models. Middlebury v3 benchmarking system is used to evaluate the error percentage and determine the accuracy. The deep learning component maintains the same default parameters as in the experiment done by [21]. Table 1 compared shelvs, jadepl and recyc images before and after the introduction of EASF. It was demonstrated that EASF generated disparity maps with a lower mean error, and the images displayed an increase in edge preservation while removing salt and pepper noise, particularly in textureless regions. The chance of a mismatch is considerable in these places since their pixel values are identical in many different places.

Table 1. Comparison of images before and after the introduction of EASF

Image	Ground truth	No EASF	With EASF
Shelvs			
Error (%)	0.0	15.0	13.5
Jadepl			
Error (%)	0.0	22.8	21.8
Recyc			
Error (%)	0.0	2.92	2.82

The quantitative evaluation findings for non-occluded error (NON-OCC) are displayed in Table 2 and for all error (ALL) in Table 3. Images of adirondack chair, motorcycle, are captured and reconstructed according to the disparity levels. The proposed method obtains the lowest error for NON-OCC error in Table 2 for the recyc and teddy images, with 2.82% and 2.92% errors, respectively. This is followed by motor at 3.94%, and motore at 3.98%. Some complex images, like shelvs and jadepl, had the lowest accuracy at 13.5% and 21.8%, respectively, due to the system's inability to recognise repeated patterns, such as the textureless regions in the shelvs image or the small repetitive leaf clusters in the jadepl image. The method outlined in this article pinpointed the exact location of the disparity, despite the fact that in certain cases it may reconstruct a disparity map with noticeable discontinuities, such as the missing mug handle in adiron or the fractured rods in ArtL. Nevertheless, the disparity level is applied precisely when the distance contours are constant and correct matching is possible. The suggested technique reduces salt-and-pepper noise while maintaining the margin separating lines. The performance of the proposed work is illustrated in Tables 2 and 3 examining it comparatively with a number of established and published methods. Table 4 compares the Middlebury ground truth to the output disparity map of the proposed system, along with the signed error for qualitative observation. The Middlebury investigation reveals that the recommended stereo corresponding approach may produce proper findings with an average non-occluded error of 6.79%. It demonstrates that the suggested method is comparable with recently published techniques and can be implemented to form an entire algorithm.

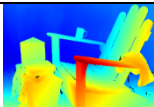
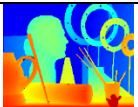
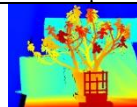
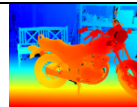
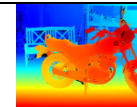
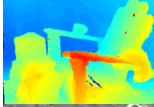
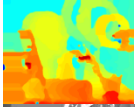
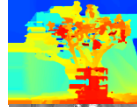
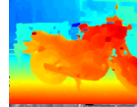
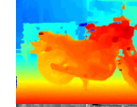
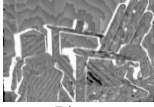



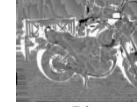

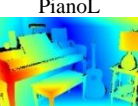

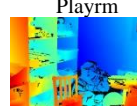
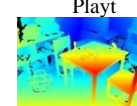
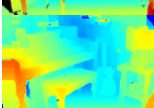
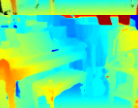
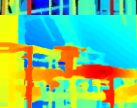
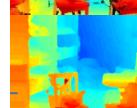
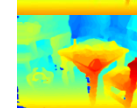
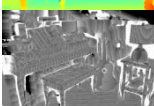
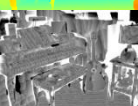
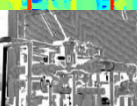
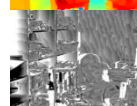
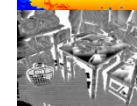
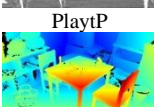
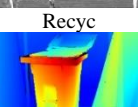


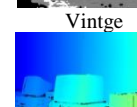
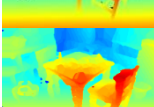
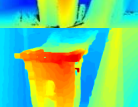
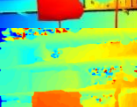
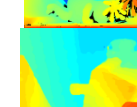
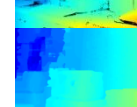

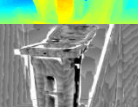
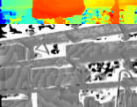
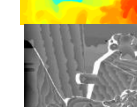
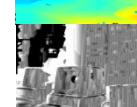
Table 2. Comparative results for NON-OCC error using Middlebury v3 evaluation platform

Method	AvgErr (%)	Adiron	ArtL	Jadepl	Motor	Piano	Pipes	Playrm	Playt	Recyc	Shelvs	Teddy	Vintge
MANet [41]	2.77	1.29	1.44	14.10	1.61	1.87	2.68	2.88	1.42	1.12	2.87	0.94	3.94
SGMEPi [42]	4.57	1.72	3.36	9.72	1.79	3.20	3.66	3.48	2.78	2.09	8.04	1.75	26.40
Proposed	6.79	4.36	7.08	21.80	3.94	4.84	7.52	4.85	6.22	2.82	13.50	2.92	8.46
PSMNet_ROB [24]	9.60	7.32	9.69	44.50	5.55	5.01	9.86	7.33	4.40	3.73	11.10	3.44	8.07
Z2ZNCC [43]	10.10	3.43	8.19	30.80	4.36	4.55	8.42	7.78	10.30	3.85	12.90	3.84	47.80

Table 3. Comparative results for ALL error using Middlebury v3 evaluation platform

Method	AvgErr (%)	Adiron	ArtL	Jadepl	Motor	Piano	Pipes	Playrm	Playt	Recyc	Shelvs	Teddy	Vintge
MANet [41]	3.30	1.48	1.75	14.90	2.11	2.13	4.95	3.83	1.67	1.28	2.99	1.23	4.69
Proposed	12.90	8.68	13.10	52.70	8.80	6.32	16.00	8.79	11.10	5.79	14.70	4.83	11.70
PSMNet_ROB [24]	13.30	8.83	13.90	68.40	8.26	5.89	14.40	9.38	5.54	4.98	11.60	3.87	9.66
SGMEPi [42]	13.40	5.65	18.20	30.80	9.18	8.49	15.80	21.00	10.70	5.80	11.00	10.70	31.90
Z2ZNCC [43]	18.00	7.65	22.30	49.30	11.70	7.96	20.10	22.80	17.70	8.04	15.40	11.60	49.80

Table 4. Ground truth of Middlebury v3 dataset compared to proposed algorithm output and error

Image	Adiron	ArtL	Jadepl	Motor	MotorE
Ground truth					
Proposed					
Signed error					
Image	Piano	PianoL	Pipes	Playrm	Playt
Ground truth					
Proposed					
Signed error					
Image	PlaytP	Recyc	Shelvs	Teddy	Vintge
Ground truth					
Proposed					
Signed error					

4. CONCLUSION

This article presents a technique for stereo matching that combines a convolutional neural network (CNN) with an edge-aware smoothing filter (EASF). CNN learn features and cost functions from dataset to produce initial disparity maps. A smoothing filter preserves depth discontinuities and smooths homogeneous zones on these maps. Stereo matching in ill-posed regions is improved by integrating the benefits of both approaches. Experimental results demonstrate the effectiveness of the proposed framework. When evaluated using Middlebury benchmarking, the system delivers accurate results with an average NON-OCC error of 6.79%. EASF is a class of nonlinear filter that can be used to smooth an image while reducing edge blurring effects such as halos and phantom edges. These filters preserve edge information while blurring an image, making them useful for improving depth estimation accuracy in stereo vision research by addressing challenges such as lighting, image complexity and pixel noise. In conclusion, this article presents a technique for stereo matching that combines CNN with EASF to improve matching accuracy in certain challenging regions. The proposed technique shows promising results and can be explored further to advance the field of stereo vision research. A more extensive discussion of the system's limitations and the nature of these

challenges is crucial. Understanding under what conditions the method may fail or perform suboptimally is essential for potential users and researchers.

ACKNOWLEDGEMENTS

The authors wish to thank Universiti Teknikal Malaysia Melaka for research grant K45002.




REFERENCES

- [1] S. Yallamandaiah and N. Purnachand, "An effective face recognition method using guided image filter and convolutional neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 3, pp. 1699–1707, 2021, doi: 10.11591/ijeecs.v23.i3.pp1699-1707.
- [2] P. Mishra and P. V. V. S. Srinivas, "Facial emotion recognition using deep convolutional neural network and smoothing, mixture filters applied during preprocessing stage," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 4, pp. 889–900, Dec. 2021, doi: 10.11591/ijai.v10.i4.pp889-900.
- [3] R. A. Hadi, L. E. George, and Z. J. Ahmed, "Automatic human ear detection approach using modified adaptive search window technique," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 19, no. 2, pp. 507–514, Apr. 2021, doi: 10.12928/TELKOMNIKA.v19i2.18320.
- [4] H. N. Abdullah and N. H. Abdulghafoor, "Objects detection and tracking using fast principle component purist and kalman filter," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 2, pp. 1317–1326, Apr. 2020, doi: 10.11591/ijece.v10i2.pp1317-1326.
- [5] W. A. Indha, N. S. Zamzam, A. Saptari, J. A. Alsayaydeh, and N. Bin Hassim, "Development of Security System Using Motion Sensor Powered by RF Energy Harvesting," in *2020 IEEE Student Conference on Research and Development, SCORd 2020*, IEEE, Sep. 2020, pp. 254–258, doi: 10.1109/SCORd50371.2020.9250984.
- [6] I. Fedorchenko, A. Oliinyk, J. A. J. Alsayaydeh, A. Kharchenko, A. Stepanenko, and V. Shkaruplyo, "Modified Genetic Algorithm To Determine the Location of the Distribution Power Supply Networks in the City," *ARNP Journal of Engineering and Applied Sciences*, vol. 15, no. 23, pp. 2850–2867, 2020.
- [7] S. M. Alturfi *et al.*, "Qualitative-based QoS performance study using hybrid ACO and PSO algorithm routing in MANET," *Journal of Physics: Conference Series*, vol. 1502, no. 1, p. 012004, Mar. 2020, doi: 10.1088/1742-6596/1502/1/012004.
- [8] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *Proceedings - IEEE Workshop on Stereo and Multi-Baseline Vision, SMBV 2001*, IEEE Comput. Soc., 2001, pp. 131–140, doi: 10.1109/SMBV.2001.988771.
- [9] R. A. Hamzah *et al.*, "A study of edge preserving filters in image matching," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 111–117, Feb. 2021, doi: 10.11591/eei.v10i1.1947.
- [10] A. Z. Jidin, R. Hussin, L. W. Fook, and M. S. Mispan, "An Automation Program for March Algorithm Fault Detection Analysis," in *2021 IEEE Asia Pacific Conference on Circuits and Systems, APCCAS 2021 and 2021 IEEE Conference on Postgraduate Research in Microelectronics and Electronics, PRIMEASIA 2021*, IEEE, Nov. 2021, pp. 149–152, doi: 10.1109/APCCAS51387.2021.9687806.
- [11] A. Z. Jidin, R. Hussin, M. S. Mispan, and L. W. Fook, "Novel March Test Algorithm Optimization Strategy for Improving Unlinked Faults Detection," in *2021 IEEE Asia Pacific Conference on Circuits and Systems, APCCAS 2021 and 2021 IEEE Conference on Postgraduate Research in Microelectronics and Electronics, PRIMEASIA 2021*, IEEE, Nov. 2021, pp. 117–120, doi: 10.1109/APCCAS51387.2021.9687791.
- [12] A. F. Kadmin, R. A. Hamzah, M. N. Abd Manap, M. S. Hamid, and S. F. Abd Gani, "Improved stereo matching algorithm based on census transform and dynamic histogram cost computation," *International Journal of Emerging Technology and Advanced Engineering*, vol. 11, no. 8, pp. 48–57, 2021, doi: 10.46338/IJETAE0821_07.
- [13] R. A. Hamzah, M. S. Hamid, A. F. Kadmin, and S. F. A. Ghani, "Improvement of stereo corresponding algorithm based on sum of absolute differences and edge preserving filter," in *Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2017*, IEEE, Sep. 2017, pp. 222–225, doi: 10.1109/ICSIPA.2017.8120610.
- [14] G. D. Hager, M. Dewan, and C. V. Stewart, "Multiple kernel tracking with SSD," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, 2004, pp. 790–797, doi: 10.1109/cvpr.2004.1315112.
- [15] R. Fan and N. Dahnoun, "Real-time implementation of stereo vision based on optimised normalised cross-correlation and propagated search range on a GPU," in *IST 2017 - IEEE International Conference on Imaging Systems and Techniques, Proceedings*, IEEE, Oct. 2017, pp. 1–6, doi: 10.1109/IST.2017.8261486.
- [16] A. Arikesh and A. K. Parvathy, "Modular multilevel inverter for renewable energy applications," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 1, pp. 1–14, Feb. 2020, doi: 10.11591/ijece.v10i1.pp1-14.
- [17] L. R. L. V. Raj, A. Jidin, C. W. M. F. Che Wan Mohd Zalani, K. Abdul Karim, G. Wee Yen, and M. H. Jopri, "Improved performance of DTC of five-phase induction machines," in *Proceedings of the 2013 IEEE 7th International Power Engineering and Optimization Conference, PEOCO 2013*, IEEE, Jun. 2013, pp. 613–618, doi: 10.1109/PEOCO.2013.6564621.
- [18] T. N. S. Tengku Zawawi, A. R. Abdullah, M. H. Jopri, T. Sutikno, N. M. Saad, and R. Sudirman, "A review of electromyography signal analysis techniques for musculoskeletal disorders," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 11, no. 3, pp. 1136–1146, Jul. 2018, doi: 10.11591/ijeecs.v11.i3.pp1136-1146.
- [19] M. Z. R. Zuber Ahmadi, A. Jidin, K. B. Jaffar, M. N. Othman, R. N. P. Nagarajan, and M. H. Jopri, "Minimization of torque ripple utilizing by 3-L CHMI in DTC," in *Proceedings of the 2013 IEEE 7th International Power Engineering and Optimization Conference, PEOCO 2013*, IEEE, Jun. 2013, pp. 636–640, doi: 10.1109/PEOCO.2013.6564625.
- [20] A. S. R. A. Subki *et al.*, "Analysis on three phase cascaded h-bridge multilevel inverter based on sinusoidal and third harmonic injected pulse width modulation via level shifted and phase shifted modulation technique," *International Journal of Power Electronics and Drive Systems*, vol. 12, no. 1, pp. 160–169, Mar. 2021, doi: 10.11591/ijpeds.v12.i1.pp160-169.
- [21] J. Žbontar and Y. Lecun, "Stereo matching by training a convolutional neural network to compare image patches," *Journal of Machine Learning Research*, vol. 17, pp. 1–32, 2016.
- [22] F. Güney and A. Geiger, "Displets: Resolving stereo ambiguities using object knowledge," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2015, pp. 4165–4175, doi: 10.1109/CVPR.2015.7299044.




- [23] A. Kendall *et al.*, “End-to-End Learning of Geometry and Context for Deep Stereo Regression,” in *Proceedings of the IEEE International Conference on Computer Vision*, IEEE, Oct. 2017, pp. 66–75, doi: 10.1109/ICCV.2017.17.
- [24] J. R. Chang and Y. S. Chen, “Pyramid Stereo Matching Network,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 5410–5418, doi: 10.1109/CVPR.2018.00567.
- [25] X. Guo, K. Yang, W. Yang, X. Wang, and H. Li, “Group-wise correlation stereo network,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2019, pp. 3268–3277, doi: 10.1109/CVPR.2019.00339.
- [26] X. Song, X. Zhao, L. Fang, H. Hu, and Y. Yu, “EdgeStereo: An Effective Multi-task Learning Network for Stereo Matching and Edge Detection,” *International Journal of Computer Vision*, vol. 128, no. 4, pp. 910–930, Apr. 2020, doi: 10.1007/s11263-019-01287-w.
- [27] X. Cheng *et al.*, “Hierarchical neural architecture search for deep stereo matching,” *Advances in Neural Information Processing Systems*, vol. 2020–December, pp. 22158–22169, 2020.
- [28] V. Tankovich, C. Häne, Y. Zhang, A. Kowdle, S. Fanello, and S. Bouaziz, “HitNet: Hierarchical Iterative Tile Refinement Network for Real-time Stereo Matching,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2021, pp. 14357–14367, doi: 10.1109/CVPR46437.2021.01413.
- [29] H. Zhao, H. Zhou, Y. Zhang, Y. Zhao, Y. Yang, and T. Ouyang, “EAI-Stereo: Error Aware Iterative Network for Stereo Matching,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13841 LNCS, 2023, pp. 3–19, doi: 10.1007/978-3-031-26319-4_1.
- [30] E. S. L. Gastal and M. M. Oliveira, “Domain Transform for Edge-Aware Image and Video Processing,” *ACM Transactions on Graphics*, vol. 30, no. 4, pp. 1–12, Jul. 2011, doi: 10.1145/2010324.1964964.
- [31] M. S. Hamid, N. A. Manap, R. A. Hamzah, A. F. Kadmin, S. F. A. Gani, and A. I. Herman, “A new function of stereo matching algorithm based on hybrid convolutional neural network,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 25, no. 1, pp. 223–231, Jan. 2022, doi: 10.11591/ijeecs.v25.i1.pp223-231.
- [32] S. F. A. Gani, R. A. Hamzah, R. Latip, S. Salam, F. Noraqillah, and A. I. Herman, “Image compression using singular value decomposition by extracting red, green, and blue channel colors,” *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 1, pp. 168–175, Feb. 2022, doi: 10.11591/eei.v11i1.2602.
- [33] J. Ni, Y. Chen, Y. Chen, J. Zhu, D. Ali, and W. Cao, “A survey on theories and applications for self-driving cars based on deep learning methods,” *Applied Sciences (Switzerland)*, vol. 10, no. 8, p. 2749, Apr. 2020, doi: 10.3390/APP10082749.
- [34] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang, “On building an accurate stereo matching system on graphics hardware,” in *Proceedings of the IEEE International Conference on Computer Vision*, IEEE, Nov. 2011, pp. 467–474, doi: 10.1109/ICCV.2011.6130280.
- [35] Y. Peng, G. Li, R. Wang, and W. Wang, “Stereo matching with space-constrained cost aggregation and segmentation-based disparity refinement,” in *Three-Dimensional Image Processing, Measurement (3DIPM), and Applications 2015*, R. Sitnik and W. Puech, Eds., Mar. 2015, p. 939309, doi: 10.1117/12.2083741.
- [36] Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu, “Constant time weighted median filtering for stereo matching and beyond,” in *Proceedings of the IEEE International Conference on Computer Vision*, IEEE, Dec. 2013, pp. 49–56, doi: 10.1109/ICCV.2013.13.
- [37] Q. Yang, “Recursive Approximation of the Bilateral Filter,” *IEEE Transactions on Image Processing*, vol. 24, no. 6, pp. 1919–1927, Jun. 2015, doi: 10.1109/TIP.2015.2403238.
- [38] M. K. Hassan, S. H. S. Ariffin, S. K. Syed Yusof, N. E. Ghazali, and M. E. A. Kanona, “Analysis of hybrid non-linear autoregressive neural network and local smoothing technique for bandwidth slice forecast,” *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 19, no. 4, pp. 1078–1089, Aug. 2021, doi: 10.12928/TELKOMNIKA.v19i4.17024.
- [39] S. Joseph, I. Hipiny, and H. Ujir, “Iban plaited mat motif classification with adaptive smoothing,” *IAES International Journal of Artificial Intelligence*, vol. 12, no. 2, pp. 840–850, Jun. 2023, doi: 10.11591/ijai.v12.i2.pp840-850.
- [40] Y. Luo, M. Marhoon, S. Al Dossary, and M. Alfaraj, “Edge-preserving smoothing and applications,” *Leading Edge (Tulsa, OK)*, vol. 21, no. 2, pp. 136–158, Feb. 2002, doi: 10.1190/1.1452603.
- [41] X. Yang, L. He, Y. Zhao, H. Sang, Z. L. Yang, and X. J. Cheng, “Multi-Attention Network for Stereo Matching,” *IEEE Access*, vol. 8, pp. 113371–113382, 2020, doi: 10.1109/ACCESS.2020.3003375.
- [42] D. Scharstein, T. Tanai, and S. N. Sinha, “Semi-global stereo matching with surface orientation priors,” in *Proceedings - 2017 International Conference on 3D Vision, 3DV 2017*, IEEE, Oct. 2018, pp. 215–224, doi: 10.1109/3DV.2017.00033.
- [43] Q. Chang *et al.*, “Efficient stereo matching on embedded GPUs with zero-means cross correlation,” *Journal of Systems Architecture*, vol. 123, p. 102366, Feb. 2022, doi: 10.1016/j.sysarc.2021.102366.

BIOGRAPHIES OF AUTHORS






Shamsul Fakhar Abd Gani    completed his Bachelor of Engineering (Computer Engineering) with honours from Universiti Malaysia Perlis and his Master of Technology (Internet and Web Computing) from Royal Melbourne Institute of Technology, Australia. Currently, he is a senior lecturer at Universiti Teknikal Malaysia Melaka conducting primarily computer vision-related research. He can be contacted at email: shamsulfakhar@utem.edu.my.






Muhammad Fahmi Miskon    is an Associate Professor at Universiti Teknikal Malaysia Melaka's Department of Mechatronics Engineering. He holds a Bachelor of Electrical Engineering (Mechatronics) from Universiti Teknologi Malaysia and a Master of Science in Mechatronics from the University of Newcastle Upon Tyne, United Kingdom. He holds a Ph.D. in Electrical and Computer Engineering from Australia's Monash University. Novelty detection and robotics are among his research interests. He can be contacted at email: fahmimiskon@utem.edu.my.






Rostam Affendi Hamzah    graduated from Universiti Teknologi Malaysia with a Bachelor of Engineering in Electronic Engineering. In 2010, he graduated from Universiti Sains Malaysia with a Master of Science in Electronic System Design Engineering. In 2017, he received Ph.D majoring in Electronic Imaging from Universiti Sains Malaysia. His research interests consist of computer vision, pattern recognition, and digital image processing. He can be contacted at email: rostamaffendi@utem.edu.my.






Mohd Saad Hamid    received his Bachelor of Engineering with a major in Computer from Multimedia University. In 2014, he received his Master of Engineering in Computer and Communication from Universiti Kebangsaan Malaysia. He was formerly employed by Telekom Malaysia Berhad and Continental Automotive Components (M) Sdn Bhd and is currently pursuing a Doctoral degree at Universiti Teknikal Malaysia Melaka. Computer vision, deep learning, image processing, and embedded systems are his research interests. He can be contacted at email: mohdsaad@utem.edu.my.



Ahmad Fauzan Kadmin    is a Chartered Engineer and Professional Technologist affiliated with UTeM as a senior lecturer with over 16 years of experience in the field of electronic and computer engineering, computer vision, and medical electronics. He graduated with a Bachelor's Degree in Electronics Engineering from Universiti Sains Malaysia and a Master's Degree in Computer and Communication Engineering from Universiti Kebangsaan Malaysia. He has published a number of technical and engineering papers in the fields of computer vision and medical image processing. He can be contacted at email: fauzan@utem.edu.my.



Adi Irwan Herman    graduated with a Bachelor's Degree in Computer Engineering Technology (Computer Systems) with honours from Universiti Teknikal Malaysia Melaka. He has served with Texas Instruments Melaka for over 7 years, and his current research interest involves computer engineering-related disciplines. He can be contacted at email: adiirwanherman@gmail.com.